The problem with time-limited, sit-down tests: Assessing language

Christa van der Walt

cvdwalt@sun.ac.za

Roadmap

1. General assessment principles
2. Assessment in language learning and teaching
3. Opportunities in remote learning and teaching contexts: using rubrics to guide performance and assessment.
4. Technological advances: Big Brother is indeed watching you!
5. Possible ways forward

General assessment principles

"A test of good quality begins its life in a deliberate design" (Weideman 2020:57).

- Questions of validity and validation.
- Questions of reliability and trustworthiness.
- What are the principles of test design?

1. Integrate the multiplicity of components of the language test so that it is a unity within that multiplicity of components, which are integrated in orderly fashion to measure a unique ability or various different, but related, sub-abilities.
2. Specify clearly to the users of the test, and where possible to the public, its circumscribed and limited scope. Exercise humility by neither overestimating, nor making inappropriate claims about, what the measurement proposed can in fact accomplish.
3. Ensure that the measurements obtained are consistent, and obtain empirical evidence for the reliability of the instrument that has been designed.
4. Ensure effective measurement by using defensibly adequate test instruments or assessment material.

5. Have an appropriately and adequately differentiated test, in which each component is organised in such a way in relation to others that it provides insight into a functionally different sub-ability, but nevertheless works together organically as a viable whole.
6. Make the test intuitively appealing, acceptable and attractive to those who take it, who use its results and who are affected by it.
7. Mount a theoretical defence of how the language ability that is being tested can best be defined, in the most current terms, or at least in terms of clearly articulated and plausible alternative theoretical paradigms or perspectives.
8. Make sure that the test yields interpretable and meaningful results; that it is intelligible and clear in all respects.

9. Ensure that the test fits the level of ability of the candidates who will take it, so that it is appropriate and has relevance for the social sphere it is intended for. Make accessible to as many as are affected by it not only the test, but also additional information about it prior to its administration, through as many and diverse media as are appropriate and feasible.

10. Ensure utility by making the test an efficient and frugal measure, and obtaining the test results efficiently to ensure that they are useful both to test takers and those who will use the results.

11. Mutually align the test with language development interventions and policies, for example with the language instruction that will either follow or precede the test, and harmonise the policy, test and instruction as closely as possible with the learning or language development foreseen in their design, and with the social environment.

12. Be prepared to give account to the users as well as to the public of how the test has been or will be used, whether its design is justifiable, and what may reasonably and legally be implied by its results.

13. Value the integrity of the test; make no compromises of quality that will undermine its status as an instrument that is fair and compassionate to everyone, and that has been designed with care and love, with the interests of the end-users in mind.

14. Spare no effort to make the test significantly trustworthy and reputable.

Time-limited and sit-down tests can be...

- Reliable
- Trustworthy

But are they valid and just?

Language production and anxiety

"For example, a student may fail a test in a language course because anxiety interfered with the learning of vocabulary items and the student thus lacks sufficient knowledge to pass the test. However, a fully competent student may also fail the same test because anxiety arousal during testing interfered with the retrieval of vocabulary items that had been mastered. Placed in a more relaxed performance context, the performance of these two students would probably be very different" (Zafar 2010:200).

To what extent do we…

"Specify clearly to the users of the test, and where possible to the public, its circumscribed and limited scope. Exercise humility by neither overestimating, nor making inappropriate claims about, what the measurement proposed can in fact accomplish (Weideman 2020:59)"?

Final examinations/ tests claim to be able to make a final judgement about students' mastery (or lack of it) in a subject field. Supplementary examinations/ tests are the worst culprits in this regard.

"Time-limited, sit-down tests."

Assessing language

1. Discrete-point assessments cannot assess language use at advanced levels (time limitation becomes a problem).

2. Traditional assessments cannot assess interaction and verbal communication (sitting down with other students becomes a problem):

"The OPI is a one-on-one criterion referenced assessment. The speaker's performance is compared to the criteria outlined in the *ACTFL Proficiency Guidelines 2012 - Speaking* or the *Interagency Language Roundtable Language Skill Level Descriptors – Speaking*. The interview is double rated and an Official ACTFL Oral proficiency Certificate stating the candidate's proficiency level is issued."

(https://www.actfl.org/assessment-research-and-development/actfl-assessments/actfl-postsecondary-assessments/oral-proficiency-interview-opi)

Different types of assessment in language teaching

Integrative v. discrete point tests

Norm-referenced and criterion-referenced

Rubrics are usually used for integrative, norm-referenced assessment.

What? And more importantly, why?

- A rubric is……
  - Also known as a marking schedule/ marking rubric.
  - A tool to ensure reliability.
  - A technology imposed on natural artefacts.
  - A limitation on creativity.
  - Prescriptive.
  - An antidote to rampant subjectivity.
  - The stupid/ inexperienced teacher's salvation.
  - Something to give students that would create the illusion of objectivity and measurability.

Open-ended, online assessments guided by a rubric

1. Complex tasks that require thinking, planning and revising.
2. More time is available (we cannot do away with time altogether).
3. Assessment criteria are stricter.
4. Students are allowed to consult with me, their peers, anybody or anything.
5. Connected more closely to real-life writing tasks.
6. Allows for integration of a variety of sources.

Using rubrics to guide performance for integrative assessment

1. The difference between a model answer, a marking rubric and a guiding rubric:

    a. A memorandum: "Students must mention at least 5 of the 8 points to get 10 marks." A 'model' answer.

    b. A marking rubric: guides the marker.

    c. A guiding rubric: guides the student.

to write 800 words from the outset: write,
duction or conclusion or sub-headings.
[20]

tion /3	Mechanics /5
aphing s the text. at flows	4, 5: Sources are acknowledged in a consistent style. No serious language errors.
is struct red raphing andom.	3: Sources are acknowledged .... but inconsistently: patterns of language errors
No phing.	Below 3: No acknowledgement, no proper system of referencing. Language errors distract from meaning of text.

20:46

Why rubrics?

- Valid assessments, but not necessarily reliable.
- Create a degree of consistency across a batch of essays or a series of talks.
- Provide an indication to the learners/ students of the focus of your assessment in open-ended tasks.

Developing a simple rubric

Two axes with the levels of proficiency and elements to be assessed:

Elements / Levels	Addressing the task	Fluency	Structure
Good			
Average			
Weak			

Developing a guiding rubric

1 Avoid technical language.

2 Make it personal.

3 Make each category explicit and distinct.

4 Discuss with students – develop with students.

Developing a guiding rubric

Be careful when naming the different levels of performance. E.g. *good, average* and *weak* could be substituted with terms like: *established, developing, emerging*, or simply the mark ranges.

Part of a rubric to assess a letter

	Addressing the task /12	Structure /3	Mechanics /5
Experien =ced /10 – 12	You discussed (1) eclecticism, (2) relating it to CAPS document(s), including the learner-focused orientation of the (3) Communicative Approach. You included references to the literature (the article, the textbook, the CAPS document(s)).	3: You used at least 3 paragraphs to structure the text.	4, 5: You acknowledged all sources in a consistent style.

But how do you know that they are doing the work themselves?

Coalescing itself within the hallmarks of pre-teen literature, *The Rest of Us Just Live Here* evokes not only a process of self-discovery but retaliates on a genre that progresses fictionalising heroic status through its cynosure of undistinguished characters seeking their own precepts of extraordinary. Rendered through the narration of Mikey, an 18 year old teenager; this young adult text voyages itself through the realms of mysticism, fantasy and realism to posit its readers in a fictionalised body from which they can relate, learn and grow. Through its copulation of pre-teen elements, this essay will address the numerous constituents underpinning the text.

Paraphrasing tools

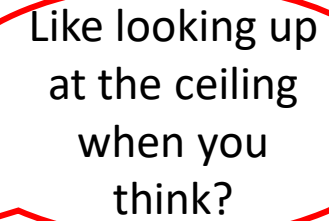Proctoring, or Big Brother is really watching you.

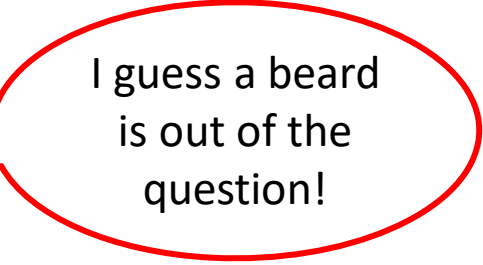Proctors in ProctorU's Hoover, Ala., office watch over students. (ProctorU)

At the start of a ProctorU test, students are told to show the proctor their student ID cards, their rooms and the tops of their desks to prove they don't have any cheating material at hand. During the test, the proctor listens through the student's microphone to ensure he or she does not ask for help from someone out of view.

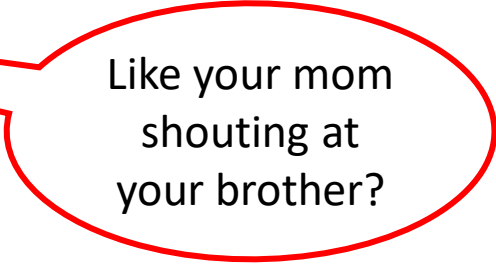Generally, online invigilation systems track the following parameters:

● anomalies in gaze direction (looking elsewhere);
● face visibility coverage
● presence of other faces in the frame
● human voice detection
● monitor switch detection
● secondary gadgets detection
● books and notes detection
● calculator detection (if not allowed)
● broadcasting software detection
● browser tabs switching detection
● third-party algorithms detection
● non-verbal communication detection

Matt (Next Level Proctoring CEO Nig... ⋯
@grandeped

I don't know if this gets enough attention, but students report that the room scans that proctoring companies require before testing include home bathrooms if the students want restroom breaks.

Think about that: they want to look at private bathrooms for "academic integrity."

Translate Tweet

23:54 · 2021/04/29 · Twitter Web App

28 Retweets **4** Quote Tweets **51** Likes

Tweet your reply

Possible ways forward

1. Move away from the point of departure that students are dishonest.

2. Implement an honour system supported by a really strict disciplinary code and serious consequences.

3. Not only flexible assessment, but also flexible contact in small groups: getting to know students.

Questions?